

Standard Normal Distribution & Confidence Intervals

Uohna Thiessen, PhD

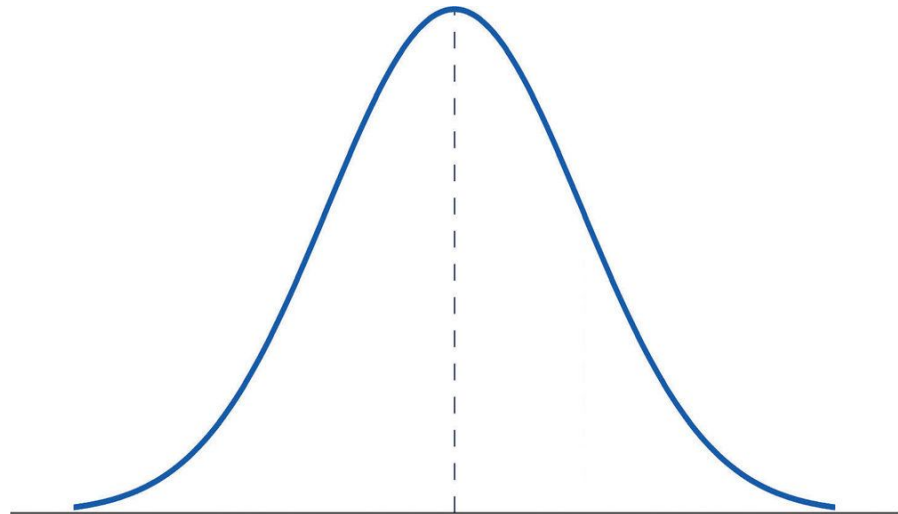
// FLATIRON SCHOOL

Outline

1. Definition of a Normal Distribution
2. Standard Normal Distribution
3. Normal Probability Functions
4. Confidence Interval
5. Comparing t & z distributions
6. Calculating Confidence Intervals

Normal/Gaussian Distribution

- **Mean = Mode = Median**
 - symmetrical, bell-shape
 - skewness = kurtosis = 0
- **described by only two parameters**
 - $N(\mu, \sigma^2)$: mean, variance
- **continuous probability distribution**
 - of many random variables
 - body temperature, IQ, heart rate



A **Normal** distribution is symmetric about the mean, values close the mean having higher probability than values far from the mean.

Standardization

Standard Normal Distribution

Data point $\rightarrow$ Mean

$$z = \frac{(x - \mu)}{\sigma}$$

σ $\rightarrow$ Standard deviation



Standardization- Comparing SAT & ACT Scores

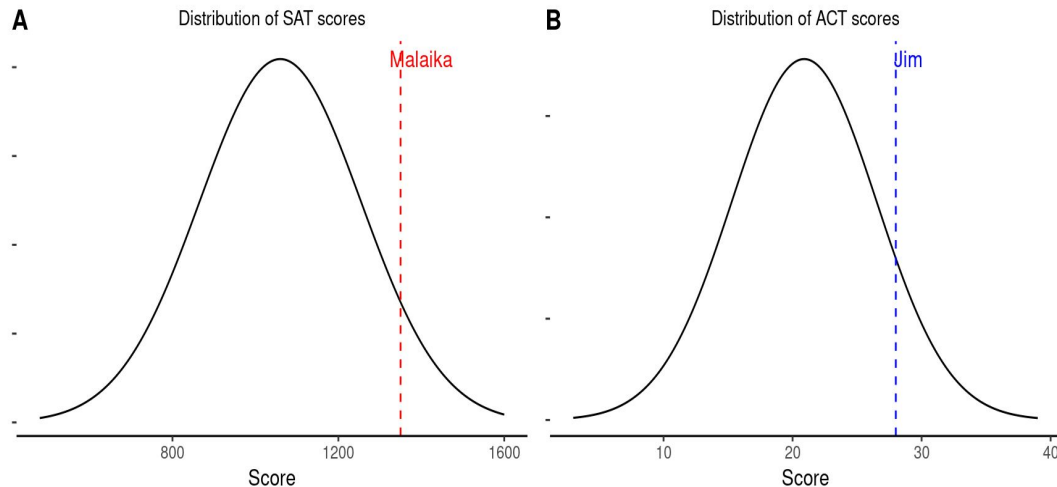
Data point $\rightarrow$ Mean

$$z = \frac{(x - \mu)}{\sigma}$$

Standard deviation $\rightarrow$

$$Z_{\text{Malaika}} = \frac{1350 - 1060}{195} = 1.49$$

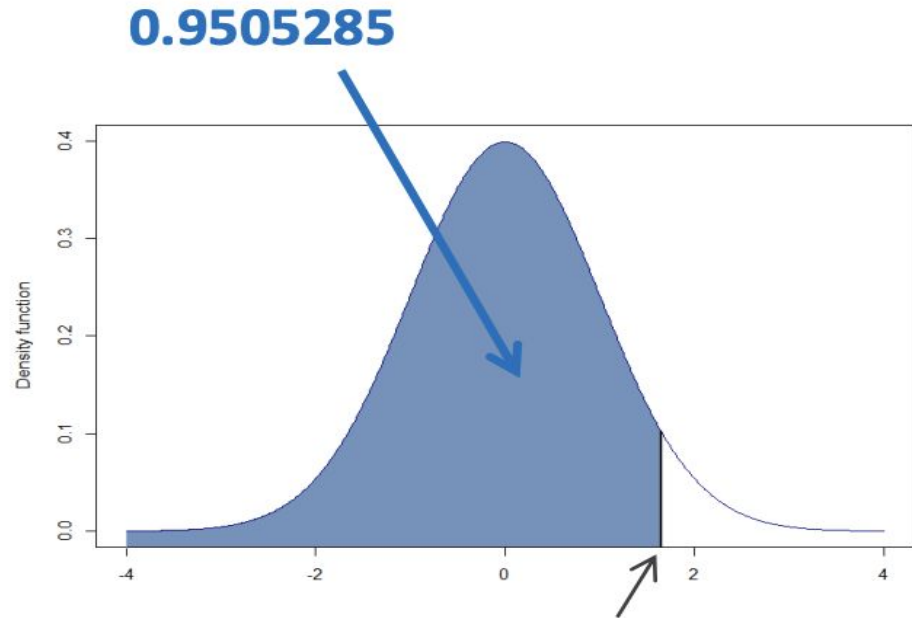
$$Z_{\text{Jim}} = \frac{28 - 20.9}{5.6} = 1.27$$



- # of standard deviations (SDs) away from the mean
- allow comparison of values with different scales

Standard Normal Distribution

```
# find the probability of Z-score of 1.65  
norm.cdf(1.65, loc = 0, scale = 1)  
>>> 0.950
```



Standard Normal Distribution

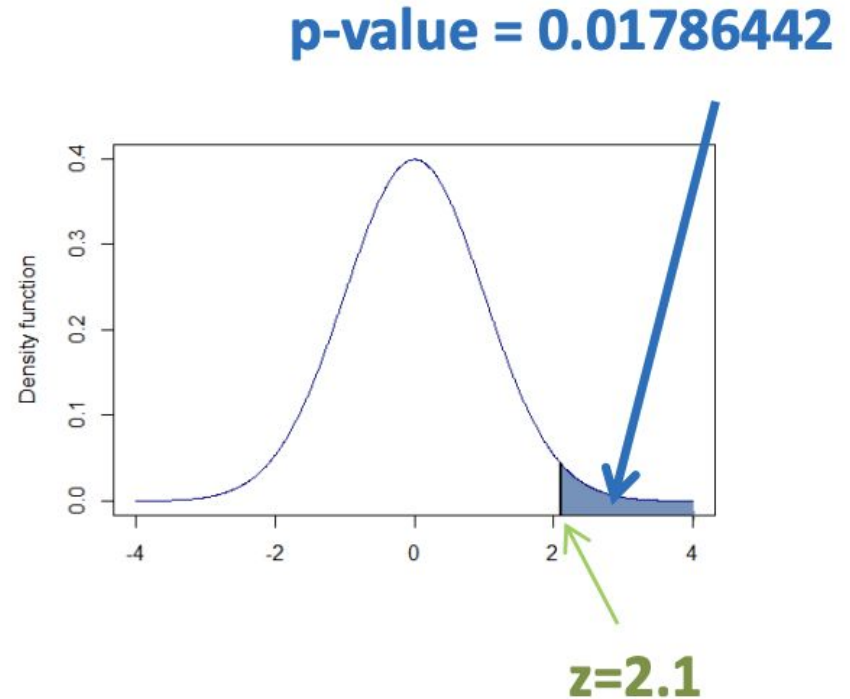
- CDF of the standard normal distribution ($\mu = 0$; $\sigma = 1$)

find the probability of Z-score of 1.65

1 - stats.norm.cdf(2.1)

stats.norm.sf(2.1)

>>> 0.0178

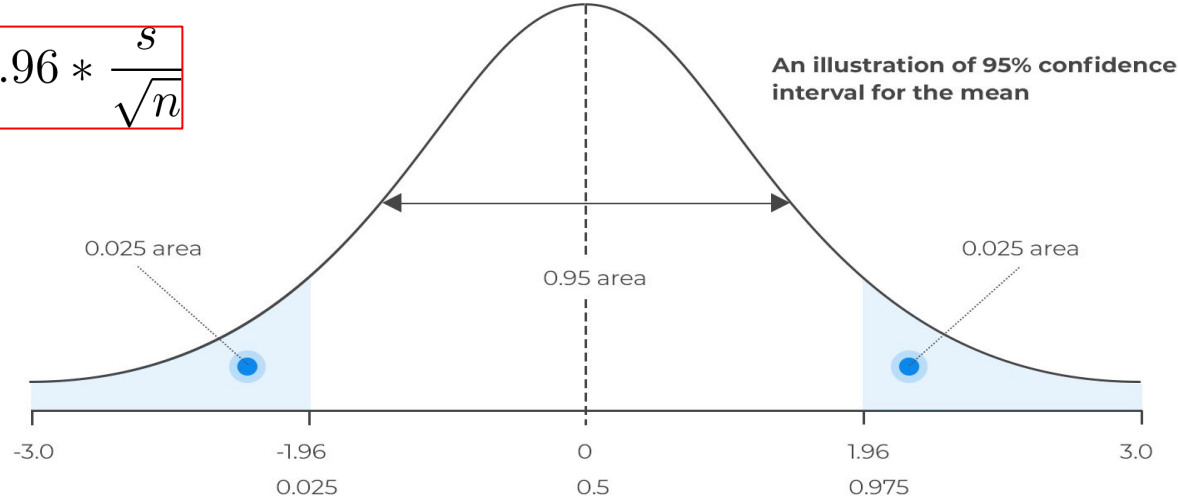


Confidence Intervals

95% Confidence Interval

95% Interval

$$CI = \mu_e \pm 1.96 * \frac{s}{\sqrt{n}}$$



Confidence Interval express the precision and uncertainty associated with a particular sampling method.

Calculating Confidence Intervals

$$\bar{x} = \frac{\sum x}{n}$$

$\bar{x}$ — mean
 $\sum x$ — data values
 n — sample size

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

σ — standard deviation

$$Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

$Z_{\alpha/2}$ — confidence coefficient
 $\alpha/2$ — confidence level
 σ — standard deviation
 n — sample size

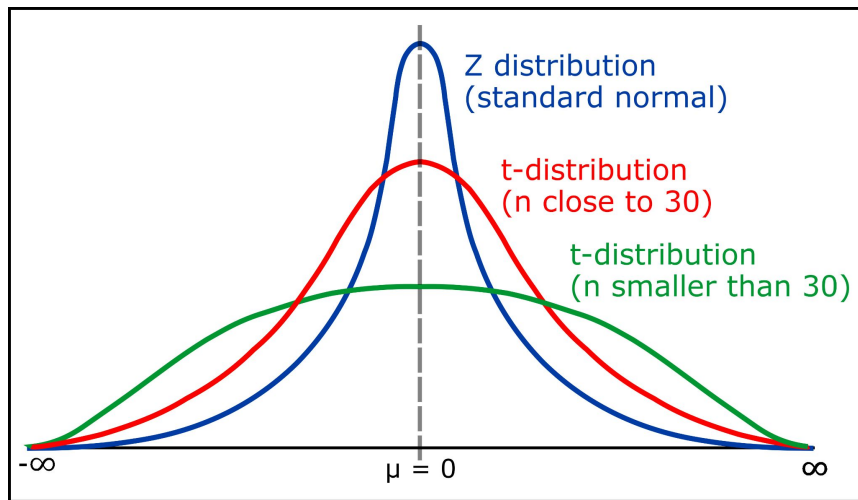
$$\bar{x} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

Comparing t- & z- Distribution

Z-score vs T-Score

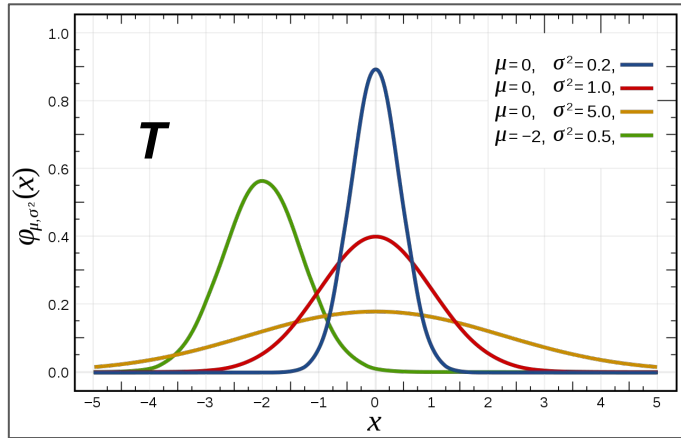
z-score : confidence interval from population mean

t-score: confidence interval from sample mean

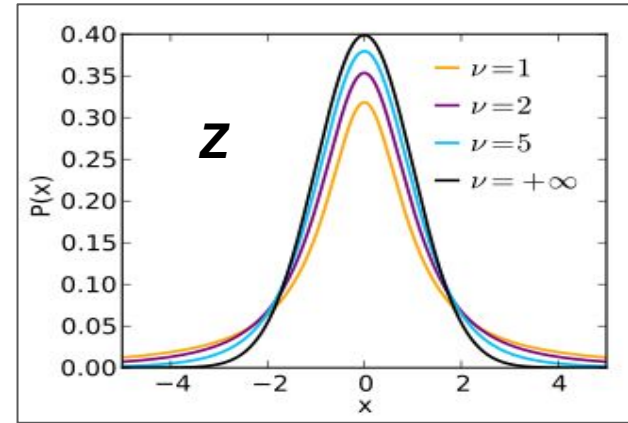


*The **t**-distribution follows the same shape as the z-distribution, but corrects for small sample sizes.
For the **t**-distribution, you need to know your degrees of freedom ($n - 1$)*

2. T- vs Z-distribution



Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 \in \mathbb{R}_{>0}$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
CDF	$\frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \right]$
Quantile	$\mu + \sigma\sqrt{2} \operatorname{erf}^{-1}(2p - 1)$
Mean	μ



Parameters	$\nu > 0$ degrees of freedom (real)
Support	$x \in (-\infty, \infty)$
PDF	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
CDF	$\frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right) \times \frac{{}_2F_1\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu} \Gamma\left(\frac{\nu}{2}\right)}$ where ${}_2F_1$ is the hypergeometric function
Mean	0 for $\nu > 1$, otherwise undefined

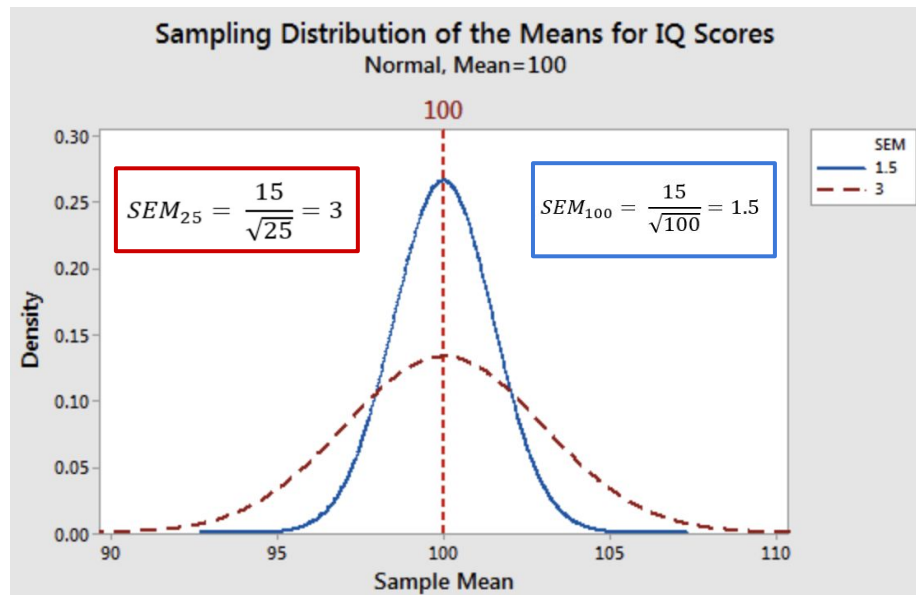
Standard Error of the Mean (SEM)

Standard Error of the Mean (SEM)

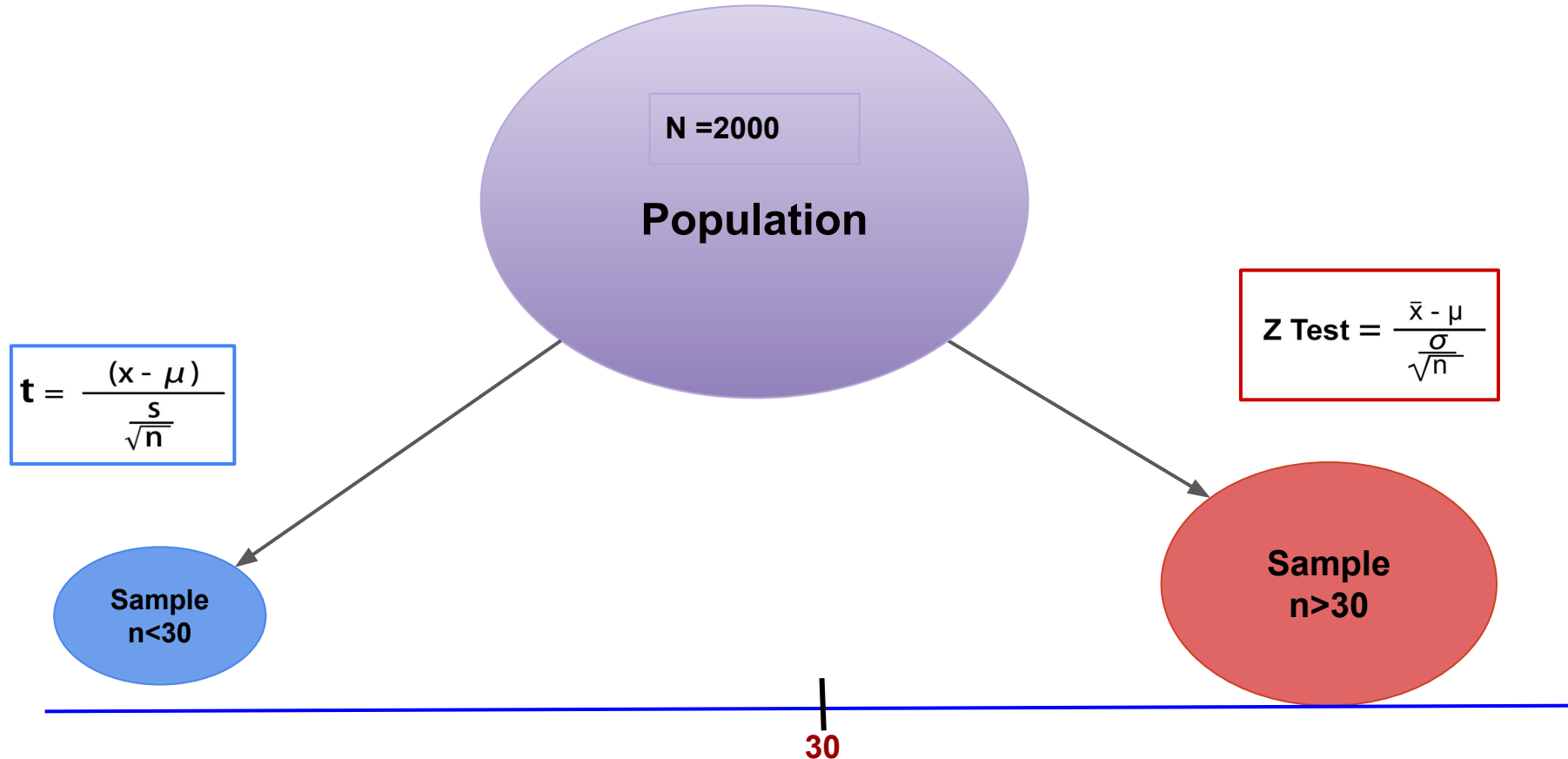
standard deviation *sample mean*

- measures of variability of the sample mean
- larger sample size (n), the smaller the **SEM**
- smaller **SEM** more precise estimation of σ

$$SEM_N = \frac{\sigma}{\sqrt{N}}$$



Z- vs. T- Distribution



Appendix

Confidence Interval Example

Example: Critical value

You survey 100 Brits and 100 Americans about their television-watching habits, and find that **both** groups watch an average of 35 hours of television per week.

The SD for the two distributions

10 for the GB estimate.

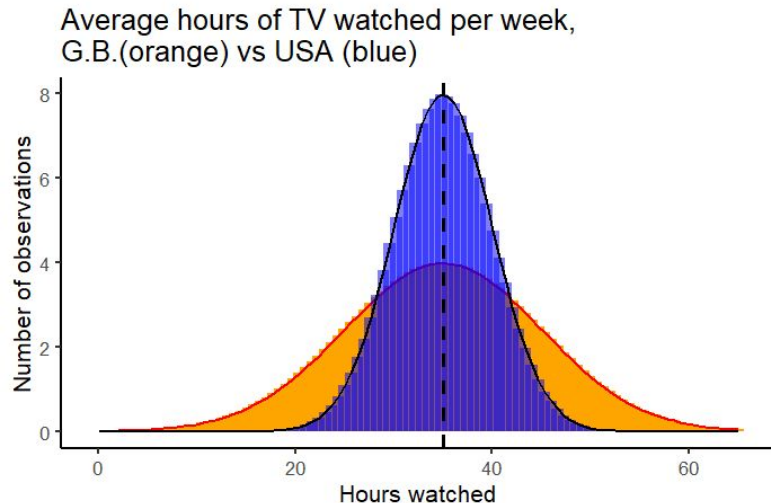
5 for the USA estimate.

$$CI = \hat{x} \pm Z^* \frac{s}{\sqrt{n}}$$

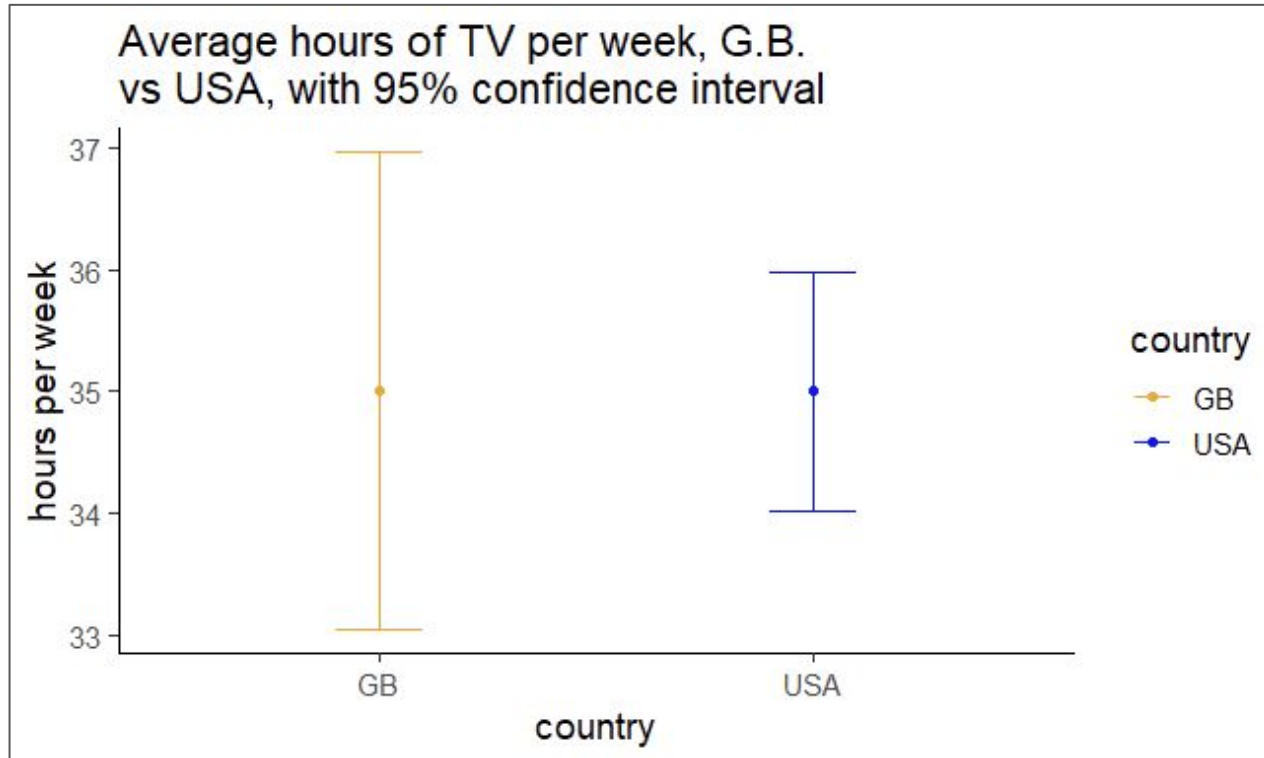
$$\begin{aligned} CI &= 35 \pm 1.96 \frac{5}{\sqrt{100}} \\ &= 35 \pm 1.96(0.5) \\ &= 35 \pm 0.98 \end{aligned}$$

$$\begin{aligned} CI &= 35 \pm 1.96 \frac{10}{\sqrt{100}} \\ &= 35 \pm 1.96(1) \\ &= 35 \pm 1.96 \end{aligned}$$

The estimate for Great Britain (**95% CI = 33.04, 36.96**)
than for the US (**95% CI = 34.02, 35.98**)



Confidence Interval Example



Probabilities Distribution

scipy.stats methods/functions

norm.pdf(): probability density function

norm.cdf(): cumulative distribution function

norm.ppf(): percent point function (inverse of cdf)

norm.sf(): survival function (1-cdf)

norm.isf(): inverse survival function (inverse of sf)

